BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 7: e112283
doi: 10.3897/biss.7.112283

OPEN ACCESS

Conference Abstract

A Novel Part in the Swiss Army Knife for Linking Biodiversity Data: The digital specimen identifier service

Wouter Addink[‡,§], Soulaine Theocharides[‡,§], Sharif Islam[§,‡]

‡ Distributed System of Scientific Collections - DiSSCo, Leiden, Netherlands
§ Naturalis Biodiversity Center, Leiden, Netherlands

Corresponding author: Wouter Addink (wouter.addink@naturalis.nl)

Received: 06 Sep 2023 | Published: 07 Sep 2023

Citation: Addink W, Theocharides S, Islam S (2023) A Novel Part in the Swiss Army Knife for Linking Biodiversity Data: The digital specimen identifier service. Biodiversity Information Science and Standards 7: e112283. https://doi.org/10.3897/biss.7.112283

Abstract

Digital specimens are new information objects on the internet, which act as digital surrogates of the physical objects they represent. They are designed to be extended with data derived from the specimen like genetic, morphological and chemical data, and with data that puts the specimen in context of its gathering event and the environment it was derived from. This requires linking the digital specimens and their related entities to information about agents, locations, publications, taxa and environmental information. To establish reliable links and (re-)connect data to specimens, a new framework is needed, which creates persistent identifiers (PIDs) for the digital specimen and its related entities. These PIDs should be actionable by machines but also can be used by humans for data citation and communication purposes.

The framework that enables this is a new PID infrastructure, produced by the European Commission-funded BiCIKL project (Biodiversity Community Integrated Knowledge Library), creates persistent and actionable identifiers. It is a generic PID infrastructure that will be used by the Distributed System for Scientific Collections research infrastructure (DiSSCo), but it can also be used by other infrastructures and institutions. PIDs minted by DiSSCo will be linked to the digital specimens and samples provided through DiSSCo. The new PIDs are a key element in enabling the concept of Digital Extended Specimens

© Addink W et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

(Webster et al. 2021) and provide unique and resolvable references to enable bidirectional linking.

DiSSCo has done extensive work to select the most appropriate PID scheme (Hardisty et al. 2021) and to design a PID infrastructure for the pan-European specimens. The draft design has been discussed with technical specialists in the joint DiSSCo and Consortium of European Taxonomic Facilities (CETAF) community, with international stakeholders like the Global Biodiversity Information Facility (GBIF) and Integrated Digitized Biocollections (iDigBio) and was discussed at the 2022 conference of the Society for the Preservation of Natural History Collections (SPNHC). A first implementation was demonstrated in the Biodiversity Information Standards (TDWG) annual conference in 2022 and illustrated key elements in the design. To be able to provide digital specimen identifiers as DOIs (Digital Object Identifiers), a pilot project was started in 2023 with DataCite to investigate if Digital Specimen DOIs in the new PID infrastructure can be created using the DataCite service. The pilot aim was to create metadata crosswalks to the DataCite schema in consultation with the DataCite Metadata Working Group, to evaluate synergies with the IGSN (International Generic Sample Number) metadata schema, to develop and test PID kernel metadata registration, and to evaluate performance and the impact of using DataCite services. There are around two billion specimens and creating PIDs for them as DOIs requires creating DOIs at an unprecedented scale. Also, PID kernel metadata registration is new for DOIs. The included metadata for specimens will complement existing Biodiversity Information Standards such as Darwin Core, and supports the new MIDS (Minimum Information about a digital specimen) standard that is under development.

The design, development and testing of the new PID infrastructure is being done as part of the BiCIKL project that aims to foster collaboration between infrastructures and develop bidirectional connections (Penev et al. 2022). In the session, we will demonstrate the results in development of the PID infrastructure as part of the BiCIKL toolbox to link biodiversity data and to discuss the progress with creating digital specimen DOIs.

Keywords

PID, BiCIKL, DiSSCo, DOI

Presenting author

Wouter Addink

Presented at

TDWG 2023

Funding program

This work has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101007492 (BiCIKL project)

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Hardisty A, Addink W, Glöckler F, Güntsch A, Islam S, Weiland C (2021) A choice of persistent identifier schemes for the Distributed System of Scientific Collections (DiSSCo). Research Ideas and Outcomes 7 https://doi.org/10.3897/rio.7.e67379
- Penev L, Koureas D, Groom Q, Lanfear J, Agosti D, Casino A, Miller J, Arvanitidis C, Cochrane G, Hobern D, Banki O, Addink W, Kõljalg U, Copas K, Mergen P, Güntsch A, Benichou L, Benito Gonzalez Lopez J, Ruch P, Martin C, Barov B, Demirova I, Hristova K (2022) Biodiversity Community Integrated Knowledge Library (BiCIKL). Research Ideas and Outcomes 8 https://doi.org/10.3897/rio.8.e81136
- Webster M, Buschbom J, Hardisty A, Bentley A (2021) The Digital Extended Specimen will Enable New Science and Applications. Biodiversity Information Science and Standards 5 https://doi.org/10.3897/biss.5.75736